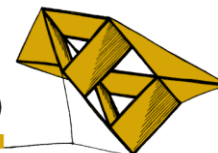




STUDENTSKÁ VĚDECKÁ KONFERENCE 2018



Modelování a odhadování výsledků tenisových zápasů

Jan Špaček¹

1 Úvod

Díky rozmachu výpočetní techniky a internetu je nyní sázení na sportovní utkání jednodušší a rychlejší než bylo dříve. Cílem sázejících je výhra. K té je potřeba vybrat „správné“ sázkové příležitosti, na které se bude sázet. Přístup při výběru sázkových příležitostí je různý. Lidé často spoléhají na svoji intuici či sází na svůj oblíbený tým. Dále lidé sází podle získaných informací, což může být například informace o aktuální formě hráče či týmu, zranění hráče atd. Další možností při výběru sázkových příležitostí je použití statistických a matematických modelů, které rovněž využívají sázkové kanceláře pro stanovení kurzů. V této práci byly modelovány a následně odhadovány výsledky tenisových zápasů na základě Elo rating systému.

2 Elo rating systém

Elo rating systém je hodnocení kvality jednotlivých hráčů a již dlouhou dobu se používá v šachu. Dá se ale využít i v jiných sportech, například v tenise. Autorem tohoto systému je Arpad Elo (1978).

Výkonnost hráče v zápase ovlivňuje mnoho náhodných veličin. V šachu to je například zdravotní stav hráče, nálada hráče, únava, doba cestování na zápas atd. V tenise to může být navíc například povrch, diváci, rozhodčí či počasí. To vedlo Arpada Elo při použití Ljapunovovy centrální limitní věty k tomu, že výkonnost hráče v zápase popsal pomocí normálního rozdělení pravděpodobnosti se střední hodnotou rovnou Elo ratingovému ohodnocení hráče R . Později bylo zjištěno, že popis výkonnosti hráče je přesnější pomocí logistického rozdělení pravděpodobnosti.

Pravděpodobnost výhry hráče A s Elo ratingem R_A nad hráčem B s Elo ratingem R_B lze vyjádřit ve tvaru

$$P_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}. \quad (1)$$

Pravděpodobnost výhry nezáleží na konkrétní hodnotě Elo ratingu, ale vždy jen na rozdílu hodnot Elo ratingů R obou soupeřů.

Lze předpokládat, že výkonnost hráče se v čase mění, tím se mění i jeho Elo rating. Proto je třeba Elo Rating hráče přepočítávat. V této práci se přepočítává Elo rating po každém odehraném zápase hráče. Nový Elo rating hráče R_N se vypočte dle vzorce

$$R_N = R_P + K \cdot (S - S_0). \quad (2)$$

K původnímu Elo ratingu hráče R_P se přičte či odečte určitý počet Elo bodů v závislosti na výsledku zápasu S , pravděpodobnosti výhry S_0 a velikosti koeficientu rozvoje K .

Nový hráč na začátku kariéry nemá odhadnutý žádný Elo rating, proto mu je třeba nějakým způsobem odhadnout úvodní Elo Rating. Zde jsou dva různé principy, jak odhadnout

¹ student navazujícího studijního programu Aplikované vědy a informatika, obor Finanční informatika a statistika, specializace Aplikovaná statistika a finance, e-mail: honza08@students.zcu.cz

úvodní Elo rating hráče. První způsob je takový, že každý nový hráč dostane stejně vysoký úvodní Elo rating. Druhou možností je, že se úvodní Elo rating hráče odhadne až po několika odehraných zápasech na základě úspěšnosti a průměrného Elo ratingu jeho soupeřů.

3 Data

K dispozici byla databáze tenisových zápasů. Celkem bylo v databázi 32 000 zápasů mužů odehraných v letech 2005 až 2016 a 24 000 zápasů žen odehraných v letech 2007 až 2016.

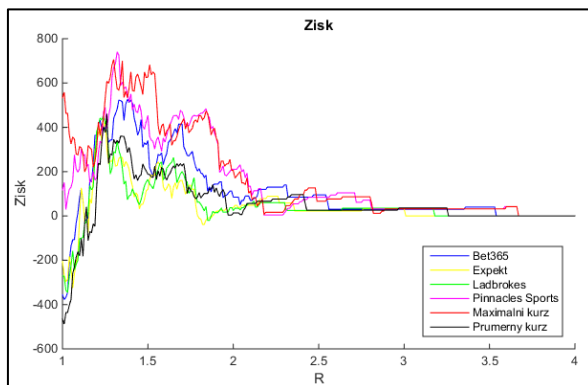
Data od roku 2005 u mužů, resp. 2007 u žen, do roku 2012 slouží k učení modelu. Na datech z let 2013 a 2014 je provedena optimalizace parametrů modelu. Po dobu těchto dvou let jsou výsledky zápasů započítávány do kritérií určujících kvalitu modelu, což jsou přesnost, logaritmická ztrátová funkce a kalibrace. Na základě těchto kritérií jsou poté parametry Elo modelu optimalizovány. Data z let 2015 a 2016 jsou použita pro ověření modelu pomocí sázení.

4 Sázení

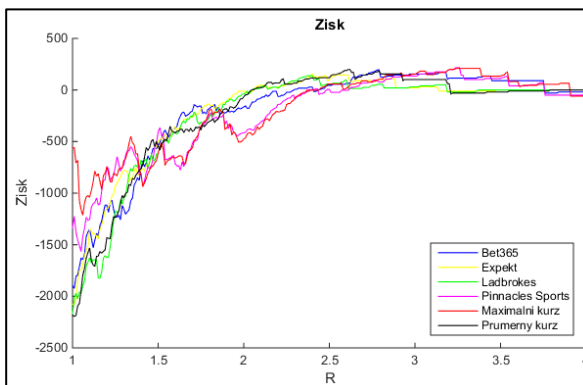
Ověření modelu bylo provedeno pomocí sázení, jak na zápasy mužů, tak na zápasy žen. Pomocí optimalizovaného Elo modelu byly spočteny pravděpodobnosti vítězství hráčů v zápasech P . Pokud pro předem stanovený koeficient sázení R byla splněna podmínka (3), tak se na zápas vsadilo 10 jednotek.

$$P \cdot \text{Kurz} \geq R \quad (3)$$

Lépe si vedl Elo model při sázení na zápasy žen. V tomto případě dosáhl model u některých sázkových kanceláří vždy zisku nezávisle na velikosti koeficientu sázení R . U mužů dosáhl model zisku jen pro vysoké hodnoty koeficientu sázení R . V tom případě však bylo sázeno jen na velmi malé množství zápasů. Na obr. 1 a obr. 2 jsou velikosti zisku v závislosti na koeficientu sázení R .



Obrázek 1: Zisk ženy



Obrázek 2: Zisk Muži

Literatura

Elo, A. (1978) *The rating of chessplayers, past and present*. New York, Arco Pub.

Kovalchik, S.A. (2016) Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12(3).